FINAL EXAM - CSC/DSC 262/462 - December 20, 2016

NAME: ____

Answer all 10 questions in the space provided. All questions have equal weight. Use the back of the page for additional space if needed (please indicate clearly where you have done this). Critical values for the standard normal, t, χ^2 and F distributions will be given as needed. The exam will last three hours. You may use a calculator, but no aid sheet is permitted.

Q1: For an *iid* sample from a normal distribution we are given sample mean $\bar{X} = 0.709$, $n = 14$, standard deviation $\sigma = 1.25$.

(a) Calculate a confidence interval for population mean μ with confidence level $1 - \alpha = 0.95$.

(b) What sample size woould be needed to obtain a margin of error $ME = 0.1$ for a confidence level $1 - \alpha = 0.99$.

[$z_{0.025} = 1.96$, $z_{0.005} = 2.587$]

Q1: [Answer]

(a) We have $\alpha = 0.05$, so we need critical value

$$z_{\alpha/2} = z_{0.025} = 1.96,$$

giving level $1 - \alpha$ confidence interval

$$\begin{aligned} CI &= \bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \\ &= 0.709 \pm 1.96 \times 0.334 \\ &= 0.709 \pm 0.655 = (0.0542, 1.364). \end{aligned}$$

(b) We have

$$n^* = \left(\frac{z_{0.01/2}\sigma}{ME} \right)^2 = \left(\frac{2.587 \times 1.25}{0.1} \right)^2 = 1045.7$$

Round up to $n^* = 1046$.

Q2: We are given two independent samples from normally distributed populations. The data is summarized in the table below.

(a) Test for equality of variances, using significance level $\alpha = 0.05$.

(b) Perform a two-sided hypothesis test using hypotheses

$$H_o: \mu_1 - \mu_2 = 0 \text{ against } H_a: \mu_1 - \mu_2 \neq 0.$$

Use significance level $\alpha = 0.05$. Use a two-sample t-test, using the conclusion of part (a) to guide your choice of method.

	Sample 1	Sample 2
$\bar{X}$	101.379	166.446
S	8.211	8.665
n	25	50

[$F_{24,49;1-0.025} = 0.474$, $F_{24,49;0.025} = 1.937$, $t_{73;0.025} = 1.993$]

Q2: [Answer]

(a) We have

$$F = S_1^2/S_2^2 = 0.898.$$

Reject $H_o: \sigma_1^2 = \sigma_2^2$ if

$$F \leq F_{1-\alpha/2,n_1-1,n_2-1} = 0.474 \text{ or } F \geq F_{\alpha/2,n_1-1,n_2-1} = 1.937$$

where $\alpha = 0.05$. Therefore, do not reject the null hypothesis of equal variances at a signficance level $\alpha = 0.05$. P-value $= 0.7941$.

(b) Use the pooled procedure with $\nu = 73$ degrees of freedom.

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{24 \times 8.211^2 + 49 \times 8.665^2}{73} = 72.56$$

or $S_p = 8.51841$. Then

$$T = \frac{\bar{X}_2 - \bar{X}_1}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{101.379 - 166.446}{8.51841\sqrt{\frac{1}{25} + \frac{1}{50}}} = -31.18363.$$

Reject H_o if

$$|T| \geq t_{\nu,\alpha/2} = 1.993.$$

Therefore, reject the null hypothesis at a signficance level $\alpha = 0.05$. P-value $= 6.184\text{e-}44$.

Q3: We are given two paired samples from normally distributed populations ($n = 6$). The data is summarized in the table below. Construct level $1 - \alpha = 0.9$ confidence interval for $\mu_1 - \mu_2$.

	Sample 1	Sample 2	Difference
1	46.993	34.846	12.147
2	44.241	45.226	-0.985
3	48.334	45.171	3.163
4	45.816	48.341	-2.525
5	46.837	39.383	7.454
6	48.748	56.129	-7.381

[$t_{5;0.05} = 2.015$]

Q3: [Answer]

$$\bar{X}_1 = 46.828, \;\; \bar{X}_2 = 44.849, \;\; \bar{X}_D = \bar{X}_1 - \bar{X}_2 = 1.979, \;\; S_D = 7.091.$$

Then the CI is given by

$$\begin{aligned} CI &= \bar{X}_D \pm t_{n-1,\alpha/2} \times \frac{S_D}{\sqrt{n}} \\ &= 1.979 \pm 2.015 \times \frac{7.091}{\sqrt{6}} \\ &= 1.979 \pm 5.834 = (-3.85, 7.813). \end{aligned}$$

Q4: Given an *iid* sample of size $n = 253$ we observe a count of $X = 46$ of a certain category. Suppose p is the population proportion of that category.

(a) Calculate a confidence interval for p with confidence level $1 - \alpha = 0.95$.

(b) What sample size is needed to ensure a margin of error $ME \leq 0.02$, assuming $p \leq 0.25$ (again, use $1 - \alpha = 0.95$)?

(c) What sample size is needed to ensure a margin of error $ME \leq 0.02$ if no assumption about p is made (again, use $1 - \alpha = 0.95$)?

$[z_{0.025} = 1.96]$

Q4: [Answer]

(a) Estimate is $\hat{p} = 46/253 = 0.1818$.

$$\begin{aligned} CI &= \hat{p} \pm z_{\alpha/2} \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\ &= 0.1818 \pm z_{\alpha/2} \times \sqrt{\frac{0.1818(1-0.1818)}{253}} \\ &= 0.1818 \pm 0.0475 \\ &= (0.134, 0.229). \end{aligned}$$

(b)
$$n^* = p^*(1-p^*)\left(\frac{z_{\alpha/2}}{ME}\right)^2 = 0.25 \times (1-0.25)\left(\frac{1.96}{0.2}\right)^2 = 1800.75.$$

Round up to $n^* = 1801$.

(c)
$$n^* = p^*(1-p^*)\left(\frac{z_{\alpha/2}}{ME}\right)^2 = 0.5 \times (1-0.5)\left(\frac{1.96}{0.2}\right)^2 = 2401.$$

Use $n^* = 2401$.

Q5: A study compares the audit rates of residents of New York state (NY) to residents of Pennsylvania (PA). The audit classifications (ie *Audit* or *No Audit*) of $n = 1621$ randomly selected study subjects, along with their state of residence, are summarized in the contingency table given below. We are interested in estimating the odds ratio $OR = Odds(Audit \mid NY)/Odds(Audit \mid PA)$. Construct a confidence interval for $\log(OR)$ with confidence level $1 - \alpha = 0.95$. Interpret the result as a two-sided test for null hypothesis $H_o : OR = 1$ against $H_a : OR \neq 1$.

	Audit	*No Audit*
NY	55	12
PA	912	642

$[z_{0.025} = 1.96]$

Q5: [Answer] The estimate of the OR is

$$\hat{OR} = \frac{n_{11}n_{22}}{n_{12}n_{21}} = \frac{55 \times 642}{12 \times 912} = 3.226.$$

The standard error of $\log(\hat{OR})$ is

$$\begin{aligned} SE(\log(\hat{OR})) &= \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}} \\ &= \sqrt{\frac{1}{55} + \frac{1}{12} + \frac{1}{912} + \frac{1}{642}} \\ &= 0.3227527. \end{aligned}$$

The CI is

$$\begin{aligned} CI &= \log(\hat{OR}) \pm z_{\alpha/2} \times SE(\log(\hat{OR})) \\ &= 1.171 \pm 1.96 \times 0.3227527 \\ &= 1.171 \pm 0.6325836 \\ &= (0.539, 1.804). \end{aligned}$$

The CI does not contain 0, equivalent to $\log(OR) = 0$ where $OR = 1$, so reject the null hypothesis at a signficance level $\alpha = 0.05$.

Q6: Consider the following *iid* sample from a normal distribution:

$$100.086, 99.954, 100.242, 99.835, 99.954, 100.083$$

with sample size $n = 6$. Calculate a confidence interval for population standard deviation σ, using confidence level $1-\alpha = 0.95$. Also give the level $1-\alpha = 0.95$ lower and upper confidence bounds.

[$\chi^2_{5;0.025} = 12.833$, $\chi^2_{5;0.975} = 0.831$, $\chi^2_{5;0.05} = 11.07$, $\chi^2_{5;0.95} = 1.145$]

Q6: [Answer] We have $S = 0.142$. The level $1 - \alpha$ confidence interval for σ is given by

$$\frac{S}{\sqrt{\left(\chi^2_{n-1,\alpha/2}\right)/(n-1)}} < \sigma < \frac{S}{\sqrt{\left(\chi^2_{n-1,1-\alpha/2}\right)/(n-1)}}.$$

The appropriate critical values are

$$\chi^2_{n-1,\alpha/2} = \chi^2_{5,0.025} = 12.833 \text{ and } \chi^2_{n-1,1-\alpha/2} = \chi^2_{5,0.975} = 0.831.$$

The confidence interval is then given by

$$\frac{0.142}{\sqrt{12.833/5}} < \sigma < \frac{0.142}{\sqrt{0.831/5}}$$

or equivalently,

$$CI = (0.0884, 0.347) = \sqrt{(0.0078, 0.1207)}.$$

The level $1 - \alpha$ lower bound for σ is given by

$$\sigma > \frac{S}{\sqrt{\left(\chi^2_{n-1,\alpha}\right)/(n-1)}}.$$

The appropriate critical value is

$$\chi^2_{n-1,\alpha} = \chi^2_{5,0.05} = 11.07.$$

The lower bound is then given by

$$\sigma > \frac{0.142}{\sqrt{11.07/5}} = 0.0952 = \sqrt{0.00906}.$$

The level $1 - \alpha$ upper bound for σ is given by

$$\sigma < \frac{S}{\sqrt{\left(\chi^2_{n-1,1-\alpha}\right)/(n-1)}}.$$

The appropriate critical value is

$$\chi^2_{n-1,1-\alpha} = \chi^2_{5,0.95} = 1.145.$$

The upper bound is then given by

$$\sigma < \frac{0.142}{\sqrt{1.145/5}} = 0.296 = \sqrt{0.0876}.$$

Q7: We are given two paired samples of sample size $n = 13$. The data is summarized in the table below. Suppose $\tilde{\mu}_D$ is the population median of the paired differences $D = X - Y$. Perform an upper tailed sign test using hypotheses $H_o : \tilde{\mu}_D = 0$ against $H_a : \tilde{\mu}_D > 0$. Use significance level $\alpha = 0.05$. Calculate a P-value using the exact binomial distribution.

	Sample 1 (X)	Sample 2 (Y)	Difference ($D = X - Y$)	Sign
1	34.1	25.6	8.5	
2	24.0	20.9	3.1	
3	37.8	36.5	1.3	
4	26.4	15.1	11.3	
5	35.5	34.5	1.0	
6	30.0	15.8	14.2	
7	25.5	19.1	6.4	
8	22.4	8.6	13.8	
9	25.5	16.7	8.8	
10	27.6	16.8	10.8	
11	34.1	21.9	12.2	
12	33.5	17.8	15.7	
13	31.6	36.1	-4.5	

Q7: [Answer] The sample median of the differences is $\tilde{D} = 8.8$. There are $X = 12$ positive differences among $n' = 13$ pairs (there are no ties). The $P-$value is

$$\begin{aligned} P &= P(X \geq 12) = P(X \leq 1) \\ &= (1-p)^{13} + np(1-p)^{12} \\ &= (1/2)^{13}(1+13) = 14/8192 = 0.0017, \end{aligned}$$

since under H_o we have $X \sim bin(13, 1/2)$, and $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$. Therefore, reject the null hypothesis at a signficance level $\alpha = 0.05$.

	Sample 1 (X)	Sample 2 (Y)	Difference ($D = X - Y$)	Sign
1	34.1	25.6	8.5	+
2	24.0	20.9	3.1	+
3	37.8	36.5	1.3	+
4	26.4	15.1	11.3	+
5	35.5	34.5	1.0	+
6	30.0	15.8	14.2	+
7	25.5	19.1	6.4	+
8	22.4	8.6	13.8	+
9	25.5	16.7	8.8	+
10	27.6	16.8	10.8	+
11	34.1	21.9	12.2	+
12	33.5	17.8	15.7	+
13	31.6	36.1	-4.5	-

Q8: We are given two paired samples of sample size $n = 9$. The data is summarized in the table below. Suppose $\tilde{\mu}_D$ is the population median of the paired differences $D = X - Y$. Perform a two-sided signed rank test using hypotheses $H_o : \tilde{\mu}_D = 0$ against $H_a : \tilde{\mu}_D \neq 0$. Use significance level $\alpha = 0.05$. Use a normal approximation without continuity correction.

	Sample 1 (X)	Sample 2 (Y)	Difference ($D = X - Y$)	Sign
1	1136.1	1125.7	10.4	
2	1264.9	1265.1	-0.2	
3	1266.4	1308.5	-42.1	
4	1160.3	1129.4	30.9	
5	1167.3	1175.4	-8.1	
6	1246.5	1184.0	62.5	
7	1125.1	1078.1	47.0	
8	1238.7	1287.5	-48.8	
9	1170.3	1235.2	-64.9	

[$z_{0.025} = 1.96$]

Q8: [Answer] The sample median of the differences is $\tilde{D} =$ -0.2. After excluding (zero) ties there are $n' = 9$ pairs remaining. The negative and positive rank sums are, respectively,

$$T_- = 1 + 2 + 5 + 7 + 9 = 24, \quad \text{and } T_+ = 3 + 4 + 6 + 8 = 21.$$

The mean and standard deviation of the negative or positive rank sums are

$$\mu_T = n(n+1)/4 = 9 \times 10/4 = 22.5,$$

and

$$\sigma_T = \sqrt{n(n+1)(2n+1)/24} = \sqrt{9 \times 10 \times 19/24} = 8.441.$$

Since $\min(T_-, T_+) = T_+$ we have Z-score

$$Z = \frac{T_+ - \mu_T}{\sigma_T} = -0.178.$$

Reject H_o if

$$|Z| \geq z_{\alpha/2} = 1.96.$$

Therefore, do not reject the null hypothesis at a signficance level $\alpha = 0.05$.

	Sample 1 (X)	Sample 2 (Y)	Difference ($D = X - Y$)	Rank $\lvert D \rvert$	Sign
1	1136.1	1125.7	10.4	3.0	+
2	1264.9	1265.1	-0.2	1.0	-
3	1266.4	1308.5	-42.1	5.0	-
4	1160.3	1129.4	30.9	4.0	+
5	1167.3	1175.4	-8.1	2.0	-
6	1246.5	1184.0	62.5	8.0	+
7	1125.1	1078.1	47.0	6.0	+
8	1238.7	1287.5	-48.8	7.0	-
9	1170.3	1235.2	-64.9	9.0	-

Q9: We are given two independent samples of sample sizes $n_1 = 5$, $n_2 = 6$. The data is summarized in the table below. Suppose $\tilde{\mu_i}$ is the population median of sample i. Perform a lower tailed rank sum test using hypotheses $H_o : \tilde{\mu_1} - \tilde{\mu_2} = 0$ against $H_a : \tilde{\mu_1} - \tilde{\mu_2} < 0$. Use significance level $\alpha = 0.05$. Use a normal approximation without continuity correction.

	1	2	3	4	5	6	$\tilde{X}_i$
Sample 1	5.6	6.3	5.5	6.6	4.6		5.6
Sample 2	7.5	7.0	6.5	7.2	6.3	6.5	6.8

[$z_{0.05} = 1.645$]

Q9: [Answer] The sample medians are $\tilde{X}_1 = 5.6$ and $\tilde{X}_2 = 6.75$. The rank sums for samples 1 and 2 are, respectively,

$$
\begin{aligned}
T_1 &= 3.0 + 4.5 + 2.0 + 8.0 + 1.0 = 18.5 \\
T_2 &= 11.0 + 9.0 + 6.5 + 10.0 + 4.5 + 6.5 = 47.5
\end{aligned}
$$

The mean and standard deviation of T_1 are

$$
\begin{aligned}
\mu_1 &= n_1(n_1 + n_2 + 1)/2 = 5 \times 12/2 = 30, \\
\sigma_W &= \sqrt{n_1 n_2 (n_1 + n_2 + 1)/12} = \sqrt{5 \times 6 \times 12/12} = \sqrt{30} = 5.477.
\end{aligned}
$$

This gives Z-score

$$
Z = \frac{T_1 - \mu_1}{\sigma_W} = \frac{18.5 - 30}{5.477} = -2.1.
$$

Reject H_o if

$$
Z \leq z_\alpha = -1.645.
$$

Therefore, reject the null hypothesis at a signficance level $\alpha = 0.05$.

	1	2	3	4	5	6	$\tilde{X}_i$
Sample 1	5.6	6.3	5.5	6.6	4.6		5.6
Sample 2	7.5	7.0	6.5	7.2	6.3	6.5	6.8
Ranks 1	3.0	4.5	2.0	8.0	1.0		18.5
Ranks 2	11.0	9.0	6.5	10.0	4.5	6.5	47.5

Q10: A sample correlation coefficient of $r = 0.62$ is observed from $n = 72$ paired observations.

(a) Let ρ be the population correlation coefficient. Test the null hypothesis $H_o : \rho = 0$ against alternative hypothesis $H_a : \rho \neq 0$ using a suitable t-statistic. Assume the samples have a bivariate normal distribution. Can we reject H_o with significance level $\alpha = 0.05$?

(b) Construct a level 95% confidence interval for ρ.

[$t_{70;0.025} = 1.994$, $z_{0.025} = 1.96$]

Q10: [Answer] The sample correlation is $r = 0.62$. We transform to T-statistic

$$T = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0.62}{\sqrt{\frac{1-0.62^2}{72-2}}} = 6.611.$$

(a) Under H_o, T has a t-distribution with $n-2 = 70$ degrees of freedom, so reject H_o at significance level α if

$$|T| \leq t_{n-2;\alpha/2} = 1.994.$$

Therefore, reject the null hypothesis at a signficance level $\alpha = 0.05$.

(b) We use transformation

$$\begin{aligned} V_{obs} &= \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) \\ &= \frac{1}{2}\ln\left(\frac{1+0.62}{1-0.62}\right) = 0.725, \end{aligned}$$

so we first use the confidence interval for $\mu_{V,\rho}$:

$$\begin{aligned} \left(V_{obs} - \frac{z_{\alpha/2}}{\sqrt{n-3}}, V_{obs} + \frac{z_{\alpha/2}}{\sqrt{n-3}}\right) &= \left(0.725 - \frac{1.96}{\sqrt{69}}, 0.725 + \frac{1.96}{\sqrt{69}}\right) \\ &= (0.489, 0.961), \end{aligned}$$

giving $L_V = 0.489$, $U_V = 0.961$. The confidence interval for ρ is then directly given by inverting the Fisher transformation,then substituting L_V and U_V:

$$\left(\frac{e^{2L_V}-1}{e^{2L_V}+1}, \frac{e^{2U_V}-1}{e^{2U_V}+1}\right) = (0.453, 0.744).$$